Estimating structural VARMA models with uncorrelated but non-independent error terms

Boubacar Mainassara, Yacouba and Francq, Christian

Université Lille III, EQUIPPE-GREMARS, Université Lille III, EQUIPPE-GREMARS

2009

Estimating structural VARMA models with uncorrelated but non-independent error terms

Y. Boubacar Mainassara[a], C. Francq[b]

[a] *Université Lille III, EQUIPPE-GREMARS, BP 60 149, 59653 Villeneuve d'Ascq cedex, France.*

[b] *Université Lille III, EQUIPPE-GREMARS, BP 60 149, 59653 Villeneuve d'Ascq cedex, France.*

Abstract

The asymptotic properties of the quasi-maximum likelihood estimator (QMLE) of vector autoregressive moving-average (VARMA) models are derived under the assumption that the errors are uncorrelated but not necessarily independent. Relaxing the independence assumption considerably extends the range of application of the VARMA models, and allows to cover linear representations of general nonlinear processes. Conditions are given for the consistency and asymptotic normality of the QMLE. A particular attention is given to the estimation of the asymptotic variance matrix, which may be very different from that obtained in the standard framework. Modified versions of the Wald, Lagrange Multiplier and Likelihood Ratio tests are proposed for testing linear restrictions on the parameters.

Key words: Echelon form, Lagrange Multiplier test, Likelihood Ratio test, Nonlinear processes, QMLE, Structural representation, VARMA models, Wald test.

1. Introduction

This paper is devoted to the problem of estimating VARMA representations of multivariate (nonlinear) processes.

Email addresses: mailto:yacouba.babacarmainassara@etu.univ-lille3.fr (Y. Boubacar Mainassara), christian.francq@univ-lille3.fr (C. Francq)

URL: http://gremars.univ-lille3.fr/ francq/ (C. Francq)

In order to give a precise definition of a linear model and of a nonlinear process, first recall that by the Wold decomposition (see *e.g.* Brockwell and Davis, 1991, for the univariate case, and Reinsel, 1997, in the multivariate framework) any zero-mean purely non deterministic d-dimensional stationary process (X_t) can be written in the form

$$X_t = \sum_{\ell=0}^{\infty} \Psi_\ell \epsilon_{t-\ell}, \qquad (\epsilon_t) \sim \mathrm{WN}(0, \Sigma) \tag{1}$$

where $\sum_\ell \|\Psi_\ell\|^2 < \infty$. The process (ϵ_t) is called the linear innovation process of the process $X = (X_t)$, and the notation $(\epsilon_t) \sim \mathrm{WN}(0, \Sigma)$ signifies that (ϵ_t) is a *weak white noise*. A weak white noise is a stationary sequence of centered and uncorrelated random variables with common variance matrix Σ. By contrast, a *strong white noise*, denoted by $\mathrm{IID}(0, \Sigma)$, is an independent and identically distributed (iid) sequence of random variables with mean 0 and variance Σ. A strong white noise is obviously a weak white noise, because independence entails uncorrelatedness, but the reverse is not true. Between weak and strong white noises, one can define a semi-strong white noise as a stationary martingale difference. An example of semi-strong white noise is the generalized autoregressive conditional heteroscedastic (GARCH) model. In the present paper, a process X is said to be *linear* when $(\epsilon_t) \sim \mathrm{IID}(0, \Sigma)$ in (1), and is said to be *nonlinear* in the opposite case. With this definition, GARCH-type processes are considered as nonlinear. Leading examples of linear processes are the VARMA and the sub-class of the vector autoregressive (VAR) models with iid noise. Nonlinear models are becoming more and more employed because numerous real time series exhibit nonlinear dynamics, for instance conditional heteroscedasticity, which can not be generated by autoregressive moving-average (ARMA) models with iid noises.[1]

The main issue with nonlinear models is that they are generally hard to identify and implement. This is why it is interesting to consider weak (V)ARMA models, that is ARMA models with weak white noises, such linear

[1]To cite few examples of nonlinear processes, let us mention the self-exciting threshold autoregressive (SETAR), the smooth transition autoregressive (STAR), the exponential autoregressive (EXPAR), the bilinear, the random coefficient autoregressive (RCA), the functional autoregressive (FAR) (see Tong, 1990, and Fan and Yao, 2003, for references on these nonlinear time series models). All these nonlinear models have been initially proposed for univariate time series, but have multivariate extensions.

representations being universal approximations of the Wold decomposition (1). Linear and nonlinear processes also have exact weak ARMA representations because a same process may satisfy several models, and many important classes of nonlinear processes admit weak ARMA representations (see Francq, Roy and Zakoïan, 2005, and the references therein).

The estimation of autoregressive moving-average (ARMA) models is however much more difficult in the multivariate than in univariate case. A first difficulty is that non trivial constraints on the parameters must be imposed for identifiability of the parameters (see Reinsel, 1997, Lütkepohl, 2005). Secondly, the implementation of standard estimation methods (for instance the Gaussian quasi-maximum likelihood estimation) is not obvious because this requires a constrained high-dimensional optimization (see Lütkepohl, 2005, for a general reference and Kascha, 2007, for a numerical comparison of alternative estimation methods of VARMA models). These technical difficulties certainly explain why VAR models are much more used than VARMA in applied works. This is also the reason why the asymptotic theory of weak ARMA model estimation is mainly limited to the univariate framework (see Francq and Zakoïan, 2005, for a review on weak ARMA models). Notable exceptions are Dufour and Pelletier (2005) who study the asymptotic properties of a generalization of the regression-based estimation method proposed by Hannan and Rissanen (1982) under weak assumptions on the innovation process, and Francq and Raïssi (2007) who study portmanteau tests for weak VAR models.

For the estimation of ARMA and VARMA models, the commonly used estimation method is the QMLE, which can also be viewed as a nonlinear least squares estimation (LSE). The asymptotic properties of the QMLE of VARMA models are well-known under the restrictive assumption that the errors ϵ_t are independent (see Lütkepohl, 2005). The asymptotic behavior of the QMLE has been studied in a much wider context by Dunsmuir and Hannan (1976) and Hannan and Deistler (1988) who proved consistency, under weak assumptions on the noise process and based on a spectral analysis. These authors also obtained asymptotic normality under a conditionally homoscedastic martingale difference assumption on the linear innovations. However, this assumption precludes most of the nonlinear models. Little is thus known when the martingale difference assumption is relaxed. Our aim in this paper is to consider a flexible VARMA specification covering the structural forms encountered in econometrics, and to relax the independence assumption, and even the martingale difference assumption, in order to be

able to cover weak VARMA representations of general nonlinear models.

The paper is organized as follows. Section 2 presents the structural weak VARMA models that we consider here. Structural forms are employed in econometrics in order to introduce instantaneous relationships between economic variables. The identifiability issues are discussed. It is shown in Section 3 that the QMLE is strongly consistent when the weak white noise (ϵ_t) is ergodic, and that the QMLE is asymptotically normally distributed when (ϵ_t) satisfies mild mixing assumptions. The asymptotic variance of the QMLE may be very different in the weak and strong cases. Section 4 is devoted to the estimation of this covariance matrix. In Section 5 it is shown how the standard Wald, LM (Lagrange multiplier) and LR (likelihood ratio) tests must be adapted in the weak VARMA case in order to test for general linearity constraints. This section is also of interest in the univariate framework because, to our knowledge, these tests have not been studied for weak ARMA models. Numerical experiments are presented in Section 6. The proofs of the main results are collected in the appendix.

2. Model and assumptions

Consider a d-dimensional stationary process (X_t) satisfying a structural VARMA(p, q) representation of the form

$$A_{00}X_t - \sum_{i=1}^{p} A_{0i}X_{t-i} = B_{00}\epsilon_t - \sum_{i=1}^{q} B_{0i}\epsilon_{t-i}, \quad (\epsilon_t) \sim \mathrm{WN}(0, \Sigma_0), \qquad (2)$$

where Σ_0 is non singular and $t \in \mathbb{Z} = \{0, \pm 1, \dots\}$. The standard VARMA$(p, q)$ form, which is sometimes called the reduced form, is obtained for $A_{00} = B_{00} = \mathrm{I}_d$. The structural forms are mainly used in econometrics to identify structural economic shocks and to allow instantaneous relationships between economic variables. Of course, constraints are necessary for the identifiability of the $(p + q + 3)d^2$ elements of the matrices involved in the VARMA equation (2). We thus assume that these matrices are parameterized by a vector ϑ_0 of lower dimension. We then write $A_{0i} = A_i(\vartheta_0)$ and $B_{0j} = B_j(\vartheta_0)$ for $i = 0, \dots, p$ and $j = 0, \dots, q$, and $\Sigma_0 = \Sigma(\vartheta_0)$, where ϑ_0 belongs to the parameter space $\Theta \subset \mathbb{R}^{k_0}$, and k_0 is the number of unknown parameters, which is typically much smaller that $(p + q + 3)d^2$. The parametrization is often linear (see Example 1 below), and thus satisfies the following smoothness conditions.

A1: The applications $\vartheta \mapsto A_i(\vartheta)$ $i = 0, \ldots, p$, $\vartheta \mapsto B_j(\vartheta)$ $j = 0, \ldots, q$ and $\vartheta \mapsto \Sigma(\vartheta)$ admit continuous third order derivatives for all $\vartheta \in \Theta$.

For simplicity we now write A_i, B_j and Σ instead of $A_i(\vartheta)$, $B_j(\vartheta)$ and $\Sigma(\vartheta)$. Let $A_\vartheta(z) = A_0 - \sum_{i=1}^p A_i z^i$ and $B_\vartheta(z) = B_0 - \sum_{i=1}^q B_i z^i$. We assume that Θ corresponds to stable and invertible representations, namely

A2: for all $\vartheta \in \Theta$, we have $\det A_\vartheta(z) \det B_\vartheta(z) \neq 0$ for all $|z| \leq 1$.

To show the strong consistency of the QMLE, we will use the following assumptions.

A3: We have $\vartheta_0 \in \Theta$, where Θ is compact.

A4: The process (ϵ_t) is stationary and ergodic.

A5: For all $\vartheta \in \Theta$ such that $\vartheta \neq \vartheta_0$, either the transfer functions

$$A_0^{-1} B_0 B_\vartheta^{-1}(z) A_\vartheta(z) \neq A_{00}^{-1} B_{00} B_{\vartheta_0}^{-1}(z) A_{\vartheta_0}(z)$$

for some $z \in \mathbb{C}$, or

$$A_0^{-1} B_0 \Sigma B_0' A_0^{-1'} \neq A_{00}^{-1} B_{00} \Sigma_0 B_{00}' A_{00}^{-1'}.$$

Remark 1. The previous identifiability assumption is satisfied when the parameter space Θ is sufficiently constrained. Note that the last condition in **A5** can be dropped for the standard reduced forms in which $A_0 = B_0 = \mathrm{I}_d$, but may be important for structural VARMA forms (see Example 1 below). The identifiability of VARMA processes has been studied in particular by Hannan (1976) who gave several procedures ensuring identifiability. In particular **A5** is satisfied when we impose $A_0 = B_0 = \mathrm{I}_d$, **A2**, the common left divisors of $A_\vartheta(L)$ and $B_\vartheta(L)$ are unimodular (*i.e.* with nonzero constant determinant), and the matrix $[A_p : B_q]$ is of full rank.

The structural form (2) allows to handle seasonal models, instantaneous economic relationships, VARMA in the so-called echelon form representation, and many other constrained VARMA representations.

Example 1. Assume that income (Inc) and consumption (Cons) variables are related by the equations $\text{Inc}_t = c_1 + \alpha_{01}\ \text{Inc}_{t-1} + \alpha_{02}\ \text{Cons}_{t-1} + \epsilon_{1t}$ and $\text{Cons}_t = c_2 + \alpha_{03}\ \text{Inc}_t + \alpha_{04}\ \text{Inc}_{t-1} + \alpha_{05}\ \text{Cons}_{t-1} + \epsilon_{2t}$. In the stationary case the process $X_t = \{\text{Inc}_t - E(\text{Inc}_t), \text{Cons}_t - E(\text{Cons}_t)\}'$ satisfies a structural VAR(1) equation

$$\begin{pmatrix} 1 & 0 \\ -\alpha_{03} & 1 \end{pmatrix} X_t = \begin{pmatrix} \alpha_{01} & \alpha_{02} \\ \alpha_{04} & \alpha_{05} \end{pmatrix} X_{t-1} + \epsilon_t.$$

We also assume that the two components of ϵ_t correspond to uncorrelated structural economic shocks, with respective variances σ_{01}^2 and σ_{02}^2. We thus have

$$\vartheta_0' = (\alpha_{01}, \alpha_{02}, \alpha_{03}, \alpha_{04}, \alpha_{05}, \sigma_{01}^2, \sigma_{02}^2).$$

Note that the identifiability condition **A5** is satisfied because for all $\vartheta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \sigma_1^2, \sigma_2^2)' \neq \vartheta_0$ we have

$$I_2 - \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_1\alpha_3 + \alpha_4 & \alpha_2\alpha_3 + \alpha_5 \end{pmatrix} z \neq I_2 - \begin{pmatrix} \alpha_{01} & \alpha_{02} \\ \alpha_{01}\alpha_{03} + \alpha_{04} & \alpha_{02}\alpha_{03} + \alpha_{05} \end{pmatrix} z$$

for some $z \in \mathbb{C}$, or

$$\begin{pmatrix} \sigma_1^2 & \sigma_1^2\alpha_3 \\ \sigma_1^2\alpha_3 & \sigma_1^2\alpha_3^2 + \sigma_2^2 \end{pmatrix} \neq \begin{pmatrix} \sigma_{01}^2 & \sigma_{01}^2\alpha_{03} \\ \sigma_{01}^2\alpha_{03} & \sigma_{01}^2\alpha_{03}^2 + \sigma_{02}^2 \end{pmatrix}.$$

3. Quasi-maximum likelihood estimation

Let $X_1, \ldots, X_n$ be observations of a process satisfying the VARMA representation (2). Note that from **A2** the matrices A_{00} and B_{00} are invertible. Introducing the innovation process

$$e_t = A_{00}^{-1} B_{00} \epsilon_t,$$

the structural representation $A_{\vartheta_0}(L)X_t = B_{\vartheta_0}(L)\epsilon_t$ can be rewritten as the reduced VARMA representation

$$X_t - \sum_{i=1}^{p} A_{00}^{-1} A_{0i} X_{t-i} = e_t - \sum_{i=1}^{q} A_{00}^{-1} B_{0i} B_{00}^{-1} A_{00} e_{t-i}. \tag{3}$$

For all $\vartheta \in \Theta$, we recursively define $\tilde{e}_t(\vartheta)$ for $t = 1, \ldots, n$ by

$$\tilde{e}_t(\vartheta) = X_t - \sum_{i=1}^{p} A_0^{-1} A_i X_{t-i} + \sum_{i=1}^{q} A_0^{-1} B_i B_0^{-1} A_0 \tilde{e}_{t-i}(\vartheta),$$

with initial values $\tilde{e}_0(\vartheta) = \cdots = \tilde{e}_{1-q}(\vartheta) = X_0 = \cdots = X_{1-p} = 0$. It will be shown that these initial values are asymptotically negligible and, in particular, that $\tilde{e}_t(\vartheta_0) - e_t \to 0$ almost surely as $t \to \infty$. The Gaussian quasi-likelihood is given by

$$\tilde{\mathrm{L}}_n(\vartheta) = \prod_{t=1}^{n} \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma_e}} \exp\left\{-\frac{1}{2}\tilde{e}_t'(\vartheta)\Sigma_e^{-1}\tilde{e}_t(\vartheta)\right\},$$

where

$$\Sigma_e = \Sigma_e(\vartheta) = A_0^{-1}B_0\Sigma B_0' A_0^{-1'}.$$

Note that the variance of e_t is $\Sigma_{e0} = \Sigma_e(\vartheta_0) = A_{00}^{-1}B_{00}\Sigma_0 B_{00}' A_{00}^{-1'}$. A quasi-maximum likelihood estimator (QMLE) is a measurable solution $\hat{\vartheta}_n$ of

$$\hat{\vartheta}_n = \arg\max_{\vartheta\in\Theta} \tilde{\mathrm{L}}_n(\vartheta) = \arg\min_{\vartheta\in\Theta} \tilde{\ell}_n(\vartheta), \quad \tilde{\ell}_n(\vartheta) = \frac{-2}{n}\log \tilde{\mathrm{L}}_n(\vartheta).$$

The following theorem shows that, for the consistency of the QMLE, the conventional assumption that the noise (ϵ_t) is an iid sequence can be replaced by the less restrictive ergodicity assumption **A4**. Dunsmuir and Hannan (1976) for VARMA in reduced form, and Hannan and Deistler (1988) for VARMAX models, obtained an equivalent result, using spectral analysis. For the proof, we do not use the spectral analysis techniques employed by the above-mentioned authors, but we follow the classical technique of Wald (1949), as was done by Rissanen and Caines (1979) to show the strong consistency of the Gaussian maximum likelihood estimator of VARMA models.

Theorem 1. *Let (X_t) be the causal solution of the VARMA equation (2) satisfying* **A1**–**A5** *and let $\hat{\vartheta}_n$ be a QMLE. Then $\hat{\vartheta}_n \to \vartheta_0$ a.s. as $n \to \infty$.*

For the asymptotic normality of the QMLE, it is necessary to assume that ϑ_0 is not on the boundary of the parameter space Θ.

A6: We have $\vartheta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ.

We now introduce mixing assumptions similar to those made by Francq and Zakoïan (1998), hereafter FZ. We denote by $\alpha_\epsilon(k)$, $k = 0, 1, \ldots$, the strong mixing coefficients of the process (ϵ_t).

A7: We have $E\|\epsilon_t\|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty}\{\alpha_\epsilon(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$.

We define the matrix of the coefficients of the reduced form (3) by

$$M_{\vartheta_0} = [A_{00}^{-1}A_{01} : \cdots : A_{00}^{-1}A_{0p} : A_{00}^{-1}B_{01}B_{00}^{-1}A_{00} : \cdots : A_{00}^{-1}B_{0q}B_{00}^{-1}A_{00} : \Sigma_{e0}].$$

Now we need an assumption which specifies how this matrix depends on the parameter ϑ_0. Let $\dot{M}_{\vartheta_0}$ be the matrix $\partial \mathrm{vec}(M_\vartheta)/\partial\vartheta'$ evaluated at ϑ_0.

A8: The matrix $\dot{M}_{\vartheta_0}$ is of full rank k_0.

One can easily verify that **A8** is satisfied in Example 1.

Theorem 2. *Under the assumptions of Theorem 1, and* **A6**-**A8**, *we have*

$$\sqrt{n}\left(\hat{\vartheta}_n - \vartheta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega := J^{-1}IJ^{-1}),$$

where $J = J(\vartheta_0)$ *and* $I = I(\vartheta_0)$, *with*

$$J(\vartheta) = \lim_{n\to\infty} \frac{\partial^2}{\partial\vartheta\partial\vartheta'}\tilde{\ell}_n(\vartheta) \quad a.s., \quad I(\vartheta) = \lim_{n\to\infty} \mathrm{Var}\,\frac{\partial}{\partial\vartheta}\tilde{\ell}_n(\vartheta).$$

For VARMA models in reduced form, it is not very restrictive to assume that the coefficients $A_0, \ldots, A_p, B_0, \ldots, B_q$ are functionally independent of the coefficient Σ_e. Thus we can write $\vartheta = (\vartheta^{(1)\prime}, \vartheta^{(2)\prime})'$, where $\vartheta^{(1)} \in \mathbb{R}^{k_1}$ depends on $A_0, \ldots, A_p$ and $B_0, \ldots, B_q$, and where $\vartheta^{(2)} \in \mathbb{R}^{k_2}$ depends on Σ_e, with $k_1 + k_2 = k_0$. With some abuse of notation, we will then write $e_t(\vartheta) = e_t(\vartheta^{(1)})$.

A9: With the previous notation $\vartheta = (\vartheta^{(1)\prime}, \vartheta^{(2)\prime})'$, where $\vartheta^{(2)} = D\,\mathrm{vec}\,\Sigma_e$ for some matrix D of size $k_2 \times d^2$.

The following theorem shows that for VARMA in reduced form, the QMLE and LSE coincide. We denote by $A\otimes B$ the Kronecker product of two matrices A and B.

Theorem 3. *Under the assumptions of Theorem 2 and* **A9** *the QMLE* $\hat{\vartheta}_n = (\hat{\vartheta}_n^{(1)\prime}, \hat{\vartheta}_n^{(2)\prime})'$ *can be obtained from*

$$\hat{\vartheta}_n^{(2)} = D\,\mathrm{vec}\,\hat{\Sigma}_e, \quad \hat{\Sigma}_e = \frac{1}{n}\sum_{t=1}^{n} \tilde{e}_t(\hat{\vartheta}_n^{(1)})\tilde{e}_t'(\hat{\vartheta}_n^{(1)}),$$

and

$$\hat{\vartheta}_n^{(1)} = \arg\min_{\vartheta^{(1)}} \det \sum_{t=1}^{n} \tilde{e}_t(\vartheta^{(1)})\tilde{e}_t'(\vartheta^{(1)}).$$

Moreover

$$J = \begin{pmatrix} J_{11} & 0 \\ 0 & J_{22} \end{pmatrix}, \quad \textit{with } J_{11} = 2E\left\{\frac{\partial}{\partial\vartheta^{(1)}} e_t'(\vartheta_0^{(1)})\right\}\Sigma_{e0}^{-1}\left\{\frac{\partial}{\partial\vartheta^{(1)'}} e_t(\vartheta_0^{(1)})\right\}$$

and $J_{22} = D(\Sigma_{e0}^{-1} \otimes \Sigma_{e0}^{-1})D'$.

Remark 2. One can see that J has the same expression in the strong and weak ARMA cases (see Lütkepohl (2005) page 480). On the contrary, the matrix I is in general much more complicated in the weak case than in the strong case.

Remark 3. In the standard strong VARMA case, *i.e.* when **A4** is replaced by the assumption that (ϵ_t) is iid, we have $I = 2J$, so that $\Omega = 2J^{-1}$. In the general case we have $I \neq 2J$. As a consequence the ready-made software used to fit VARMA do not provide a correct estimation of Ω for weak VARMA processes. The problem also holds in the univariate case (see Francq and Zakoïan, 2007, and the references therein).

4. Estimating the asymptotic variance matrix

Theorem 2 can be used to obtain confidence intervals and significance tests for the parameters. The asymptotic variance Ω must however be estimated. The matrix J can easily be estimated by its empirical counterpart. For instance, under **A9**, one can take

$$\hat{J} = \begin{pmatrix} \hat{J}_{11} & 0 \\ 0 & \hat{J}_{22} \end{pmatrix}, \quad \hat{J}_{11} = \frac{2}{n}\sum_{t=1}^{n}\left\{\frac{\partial}{\partial\vartheta^{(1)}}\tilde{e}_t'(\hat{\vartheta}_n^{(1)})\right\}\hat{\Sigma}_e^{-1}\left\{\frac{\partial}{\partial\vartheta^{(1)'}}\tilde{e}_t(\hat{\vartheta}_n^{(1)})\right\},$$

and $\hat{J}_{22} = D(\hat{\Sigma}_e^{-1} \otimes \hat{\Sigma}_e^{-1})D'$. In the standard strong VARMA case $\hat{\Omega} = 2\hat{J}^{-1}$ is a strongly consistent estimator of Ω. In the general weak VARMA case this estimator is not consistent when $I \neq 2J$ (see Remark 3). So we need a consistent estimator of I. Note that

$$I = \mathrm{Var}_{as}\,\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Upsilon_t = \sum_{h=-\infty}^{+\infty}\mathrm{Cov}(\Upsilon_t, \Upsilon_{t-h}), \tag{4}$$

where

$$\Upsilon_t = \frac{\partial}{\partial \vartheta} \left\{ \log \det \Sigma_e + e_t'(\vartheta^{(1)}) \Sigma_e^{-1} e_t(\vartheta^{(1)}) \right\}_{\vartheta=\vartheta_0}. \tag{5}$$

In the econometric literature the nonparametric kernel estimator, also called heteroscedastic autocorrelation consistent (HAC) estimator (see Newey and West, 1987, or Andrews, 1991), is widely used to estimate covariance matrices of the form I. Let $\hat{\Upsilon}_t$ be the vector obtained by replacing ϑ_0 by $\hat{\vartheta}_n$ in Υ_t. The matrix Ω is then estimated by a "sandwich" estimator of the form

$$\hat{\Omega}^{\mathrm{HAC}} = \hat{J}^{-1} \hat{I}^{\mathrm{HAC}} \hat{J}^{-1}, \quad \hat{I}^{\mathrm{HAC}} = \frac{1}{n} \sum_{t,s=1}^{n} \omega_{|t-s|} \hat{\Upsilon}_t \hat{\Upsilon}_s,$$

where $\omega_0, \ldots, \omega_{n-1}$ is a sequence of weights (see Andrews, 1991, and Newey and West, 1987, for the problem of the choice of weights).

Interpreting $(2\pi)^{-1} I$ as the spectral density of the stationary process (Υ_t) evaluated at frequency 0 (see Brockwell and Davis, 1991, p. 459), an alternative method consists in using a parametric AR estimate of the spectral density of (Υ_t). This approach, which has been studied by Berk (1974) (see also den Haan and Levin, 1997), rests on the expression

$$I = \mathbf{\Phi}^{-1}(1) \Sigma_u \mathbf{\Phi}^{-1}(1)$$

when (Υ_t) satisfies an AR(∞) representation of the form

$$\mathbf{\Phi}(L)\Upsilon_t := \Upsilon_t + \sum_{i=1}^{\infty} \Phi_i \Upsilon_{t-i} = u_t, \tag{6}$$

where u_t is a weak white noise with variance matrix Σ_u. Let $\hat{\mathbf{\Phi}}_r(z) = \mathrm{I}_{k_0} + \sum_{i=1}^{r} \hat{\Phi}_{r,i} z^i$, where $\hat{\Phi}_{r,1}, \cdots, \hat{\Phi}_{r,r}$ denote the coefficients of the LS regression of $\hat{\Upsilon}_t$ on $\hat{\Upsilon}_{t-1}, \cdots, \hat{\Upsilon}_{t-r}$. Let $\hat{u}_{r,t}$ be the residuals of this regression, and let $\hat{\Sigma}_{\hat{u}_r}$ be the empirical variance of $\hat{u}_{r,1}, \ldots, \hat{u}_{r,n}$.

We are now able to state the following theorem, which is an extension of a result given in Francq, Roy and Zakoïan (2005).

Theorem 4. *In addition to the assumptions of Theorem 2, assume that the process (Υ_t) defined in (5) admits an AR(∞) representation (6) in which the roots of $\det \mathbf{\Phi}(z) = 0$ are outside the unit disk, $\|\Phi_i\| = o(i^{-2})$, and $\Sigma_u = Var(u_t)$ is non-singular. Moreover we assume that $\|\epsilon_t\|_{8+4\nu} < \infty$ and*

$\sum_{k=0}^{\infty}\{\alpha_{X,\epsilon}(k)\}^{\nu/(2+\nu)} < \infty$ *for some* $\nu > 0$, *where* $\{\alpha_{X,\epsilon}(k)\}_{k\geq 0}$ *denotes the sequence of the strong mixing coefficients of the process* $(X_t', \epsilon_t')'$. *Then the spectral estimator of* I

$$\hat{I}^{\mathrm{SP}} := \hat{\mathbf{\Phi}}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\mathbf{\Phi}}_r'^{-1}(1) \to I$$

in probability when $r = r(n) \to \infty$ *and* $r^3/n \to 0$ *as* $n \to \infty$.

5. Testing linear restrictions on the parameter

It may be of interest to test s_0 linear constraints on the elements of ϑ_0 (in particular $A_{0p} = 0$ or $B_{0q} = 0$). We thus consider a null hypothesis of the form

$$H_0 : R_0\vartheta_0 = \mathfrak{r}_0$$

where R_0 is a known $s_0 \times k_0$ matrix of rank s_0 and $\mathfrak{r}_0$ is a known s_0-dimensional vector. The Wald, LM and LR principles are employed frequently for testing H_0. The LM test is also called the score or Rao-score test. We now examine if these principles remain valid in the non standard framework of weak VARMA models.

Let $\hat{\Omega} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}$, where $\hat{J}$ and $\hat{I}$ are consistent estimator of J and I, as defined in Section 4. Under the assumptions of Theorems 2 and 4, and the assumption that I is invertible, the Wald statistic

$$\mathbf{W}_n = n(R_0\hat{\vartheta}_n - \mathfrak{r}_0)'(R_0\hat{\Omega}R_0')^{-1}(R_0\hat{\vartheta}_n - \mathfrak{r}_0)$$

asymptotically follows a $\chi^2_{s_0}$ distribution under H_0. Therefore, the standard formulation of the Wald test remains valid. More precisely, at the asymptotic level α, the Wald test consists in rejecting H_0 when $\mathbf{W}_n > \chi^2_{s_0}(1-\alpha)$. It is however important to note that a consistent estimator of the form $\hat{\Omega} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}$ is required. The estimator $\hat{\Omega} = 2\hat{J}^{-1}$, which is routinely used in the time series softwares, is only valid in the strong VARMA case.

We now turn to the LM test. Let $\hat{\vartheta}_n^c$ be the restricted QMLE of the parameter under H_0. Define the Lagrangean

$$\mathcal{L}(\vartheta, \lambda) = \tilde{\ell}_n(\vartheta) - \lambda'(R_0\vartheta - \mathfrak{r}_0),$$

where λ denotes a s_0-dimensional vector of Lagrange multipliers. The first-order conditions yield

$$\frac{\partial\tilde{\ell}_n}{\partial\vartheta}(\hat{\vartheta}_n^c) = R_0'\hat{\lambda}, \quad R_0\hat{\vartheta}_n^c = \mathfrak{r}_0.$$

It will be convenient to write $a \overset{c}{=} b$ to signify $a = b + c$. A Taylor expansion gives under H_0

$$0 \;=\; \sqrt{n}\frac{\partial \tilde{\ell}_n(\hat{\vartheta}_n)}{\partial \vartheta} \overset{o_P(1)}{=} \sqrt{n}\frac{\partial \tilde{\ell}_n(\hat{\vartheta}_n^c)}{\partial \vartheta} - J\sqrt{n}\left(\hat{\vartheta}_n - \hat{\vartheta}_n^c\right).$$

We deduce that

$$\sqrt{n}(R_0\hat{\vartheta}_n - \mathfrak{r}_0) \;=\; R_0\sqrt{n}(\hat{\vartheta}_n - \hat{\vartheta}_n^c) \overset{o_P(1)}{=} R_0 J^{-1}\sqrt{n}\frac{\partial \tilde{\ell}_n(\hat{\vartheta}_n^c)}{\partial \vartheta} = R_0 J^{-1} R_0' \sqrt{n}\hat{\lambda}.$$

Thus under H_0 and the previous assumptions,

$$\sqrt{n}\hat{\lambda} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, (R_0 J^{-1} R_0')^{-1} R_0 \Omega R_0' (R_0 J^{-1} R_0')^{-1}\right\}, \tag{7}$$

so that the LM statistic is defined by

$$\begin{aligned}\mathbf{LM}_n &= n\hat{\lambda}'\left\{(R_0\hat{J}^{-1}R_0')^{-1}R_0\hat{\Omega}R_0'(R_0\hat{J}^{-1}R_0')^{-1}\right\}^{-1}\hat{\lambda} \\ &= n\frac{\partial \tilde{\ell}_n}{\partial \vartheta'}(\hat{\vartheta}_n^c)\hat{J}^{-1}R_0'\left(R_0\hat{\Omega}R_0'\right)^{-1}R_0\hat{J}^{-1}\frac{\partial \tilde{\ell}_n}{\partial \vartheta}(\hat{\vartheta}_n^c).\end{aligned}$$

Note that in the strong VARMA case, $\hat{\Omega} = 2\hat{J}^{-1}$ and the LM statistic takes the more conventional form $\mathbf{LM}_n^* = (n/2)\hat{\lambda}' R_0 \hat{J}^{-1} R_0' \hat{\lambda}$. In the general case, strong and weak as well, the convergence (7) implies that the asymptotic distribution of the $\mathbf{LM}_n$ statistic is $\chi^2_{s_0}$ under H_0. The null is therefore rejected when $\mathbf{LM}_n > \chi^2_{s_0}(1-\alpha)$. Of course the conventional LM test with rejection region $\mathbf{LM}_n^* > \chi^2_{s_0}(1-\alpha)$ is not asymptotically valid for general weak VARMA models.

Standard Taylor expansions show that

$$\sqrt{n}(\hat{\vartheta}_n - \hat{\vartheta}_n^c) \overset{o_P(1)}{=} -\sqrt{n}J^{-1}R_0'\hat{\lambda},$$

and that the LR statistic satisfies

$$\mathbf{LR}_n := 2\left\{\log \tilde{\mathrm{L}}_n(\hat{\vartheta}_n) - \log \tilde{\mathrm{L}}_n(\hat{\vartheta}_n^c)\right\} \overset{o_P(1)}{=} \frac{n}{2}(\hat{\vartheta}_n - \hat{\vartheta}_n^c)' J (\hat{\vartheta}_n - \hat{\vartheta}_n^c) \overset{o_P(1)}{=} \mathbf{LM}_n^*.$$

Using the previous computations and standard results on quadratic forms of normal vectors (see *e.g.* Lemma 17.1 in van der Vaart, 1998), we find that

the $\mathbf{LR}_n$ statistic is asymptotically distributed as $\sum_{i=1}^{s_0} \lambda_i Z_i^2$ where the Z_i's are iid $\mathcal{N}(0,1)$ and $\lambda_1, \ldots, \lambda_{s_0}$ are the eigenvalues of

$$\Sigma_{\mathbf{LR}} = J^{-1/2} S_{\mathbf{LR}} J^{-1/2}, \quad S_{\mathbf{LR}} = \frac{1}{2} R_0' (R_0 J^{-1} R_0')^{-1} R_0 \Omega R_0' (R_0 J^{-1} R_0')^{-1} R_0.$$

Note that when $\Omega = 2J^{-1}$, the matrix $\Sigma_{\mathbf{LR}} = J^{-1/2} R_0' (R_0 J^{-1} R_0')^{-1} R_0 J^{-1/2}$ is a projection matrix. Its eigenvalues are therefore equal to 0 and 1, and the number of eigenvalues equal to 1 is $\operatorname{Tr} J^{-1/2} R_0' (R_0 J^{-1} R_0')^{-1} R_0 J^{-1/2} = \operatorname{Tr} I_{s_0} = s_0$. Therefore we retrieve the well-known result that $\mathbf{LR}_n \sim \chi^2_{s_0}$ under H_0 in the strong VARMA case. In the weak VARMA case, the asymptotic null distribution of $\mathbf{LR}_n$ is complicated. It is possible to evaluate the distribution of a quadratic form of a Gaussian vector by means of the Imhof algorithm (Imhof, 1961), but the algorithm is time consuming. An alternative is to use the transformed statistic

$$\frac{n}{2} (\hat{\vartheta}_n - \hat{\vartheta}_n^c)' \hat{J} \hat{S}_{\mathbf{LR}}^- \hat{J} (\hat{\vartheta}_n - \hat{\vartheta}_n^c) \tag{8}$$

which follows a $\chi^2_{s_0}$ under H_0, when $\hat{J}$ and $\hat{S}_{\mathbf{LR}}^-$ are weakly consistent estimators of J and of a generalized inverse of $S_{\mathbf{LR}}$. The estimator $\hat{S}_{\mathbf{LR}}^-$ can be obtained from the singular value decomposition of any weakly consistent estimator $\hat{S}_{\mathbf{LR}}$ of $S_{\mathbf{LR}}$. More precisely, defining the diagonal matrix $\hat{\Lambda} = \operatorname{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_{k_0})$ where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_{k_0}$ denote the eigenvalues of the symmetric matrix $\hat{S}_{\mathbf{LR}}$, and denoting by $\hat{P}$ an orthonormal matrix such that $\hat{S}_{\mathbf{LR}} = \hat{P} \hat{\Lambda} \hat{P}'$, one can set

$$\hat{S}_{\mathbf{LR}}^- = \hat{P} \hat{\Lambda}^- \hat{P}', \quad \hat{\Lambda}^- = \operatorname{diag}\left(\hat{\lambda}_1^{-1}, \ldots, \hat{\lambda}_{s_0}^{-1}, 0, \ldots, 0\right).$$

The matrix $\hat{S}_{\mathbf{LR}}^-$ then converges weakly to a matrix $S_{\mathbf{LR}}^-$ satisfying $S_{\mathbf{LR}} S_{\mathbf{LR}}^- S_{\mathbf{LR}} = S_{\mathbf{LR}}$, because $S_{\mathbf{LR}}$ has full rank s_0.

6. Numerical illustrations

We first study numerically the behaviour of the QMLE for strong and weak VARMA models of the form

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & a_1(2,2) \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ b_1(2,1) & b_1(2,2) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix}, \tag{9}$$

where

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \text{IID}\,\mathcal{N}(0, I_2), \tag{10}$$

in the strong case, and

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}(|\eta_{1,t-1}| + 1)^{-1} \\ \eta_{2,t}(|\eta_{2,t-1}| + 1)^{-1} \end{pmatrix}, \quad \text{with } \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \text{IID}\,\mathcal{N}(0, I_2), \tag{11}$$

in the weak case. Model (9) is a VARMA(1,1) model in echelon form. The noise defined by (11) is a direct extension of a weak noise defined by Romano and Thombs (1996) in the univariate case. The numerical illustrations of this section are made with the free statistical software R (see http://cran.r-project.org/). We simulated $N = 1,000$ independent trajectories of size $n = 2,000$ of Model (9), first with the strong Gaussian noise (10), second with the weak noise (11). Figure 1 compares the distribution of the QMLE in the strong and weak noise cases. The distributions of $\hat{a}_1(2,2)$ and $\hat{b}_1(2,1)$ are similar in the two cases, whereas the QMLE of $\hat{b}_1(2,2)$ is more accurate in the weak case than in the strong one. Similar simulation experiments, not reported here, reveal that the situation is opposite, that is the QMLE is more accurate in the strong case than in the weak case, when the weak noise is defined by $\epsilon_{i,t} = \eta_{i,t}\eta_{i,t-1}$ for $i = 1, 2$. This is in accordance with the results of Romano and Thombs (1996) who showed that, with similar noises, the asymptotic variance of the sample autocorrelations can be greater or less than 1 as well (1 is the asymptotic variance for strong white noises).

Figure 2 compares the standard estimator $\hat{\Omega} = 2\hat{J}^{-1}$ and the sandwich estimator $\hat{\Omega} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}$ of the QMLE asymptotic variance Ω. We used the spectral estimator $\hat{I} = \hat{I}^{\text{SP}}$ defined in Theorem 4, and the AR order r is automatically selected by AIC, using the function `VARselect()` of the **vars** R package. In the strong VARMA case we know that the two estimators are consistent. In view of the two top panels of Figure 2, it seems that the sandwich estimator is less accurate in the strong case. This is not surprising because the sandwich estimator is more robust, in the sense that this estimator continues to be consistent in the weak VARMA case, contrary to the standard estimator. It is clear that in the weak case $n\text{Var}\left\{\hat{b}_1(2,2) - b_1(2,2)\right\}^2$ is better estimated by $\hat{\Omega}^{\text{SP}}(3,3)$ (see the box-plot (c) of the right-bottom panel of Figure 2) than by $2\hat{J}^{-1}(3,3)$ (box-plot (c) of the left-bottom panel). The failure of the standard estimator of Ω in the weak VARMA framework may have important consequences in terms of identification or hypothesis testing.

Table 1 displays the empirical sizes of the standard Wald, LM and LR tests, and that of the modified versions proposed in Section 5. For the nominal level $\alpha = 5\%$, the empirical size over the $N = 1,000$ independent replications should vary between the significant limits 3.6% and 6.4% with probability 95%. For the nominal level $\alpha = 1\%$, the significant limits are 0.3% and 1.7%, and for the nominal level $\alpha = 10\%$, they are 8.1% and 11.9%. When the relative rejection frequencies are outside the significant limits, they are displayed in bold type in Table 1. For the strong VARMA model I, all the relative rejection frequencies are inside the significant limits. For the weak VARMA model II, the relative rejection frequencies of the standard tests are definitely outside the significant limits. Thus the error of first kind is well controlled by all the tests in the strong case, but only by modified versions of the tests in the weak case. Table 2 shows that the powers of all the tests are very similar in the Model III case. The same is also true for the two modified tests in the Model IV case. The empirical powers of the standard tests are hardly interpretable for Model IV, because we have already seen in Table 1 that the standard versions of the tests do not well control the error of first kind in the weak VARMA framework.

From these simulation experiments and from the asymptotic theory, we draw the conclusion that the standard methodology, based on the QMLE, allows to fit VARMA representations of a wide class of nonlinear multivariate time series. This standard methodology, including in particular the significance tests on the parameters, needs however to be adapted to take into account the possible lack of independence of the errors terms. In future works, we intent to study how the existing identification (see *e.g.* Nsiri and Roy, 1996) and diagnostic checking (see *e.g.* Duchesne and Roy, 2004) procedures should be adapted in the weak VARMA framework considered in the present paper.

A. Technical proofs

We begin with a lemma useful to show the identifiability of ϑ_0.

Lemma 1. *Assume that Σ_0 is non singular and that* **A5** *holds true. If $A_0^{-1}B_0B_\vartheta^{-1}(L)A_\vartheta(L)X_t = A_{00}^{-1}B_{00}\epsilon_t$ with probability one and $A_0^{-1}B_0\Sigma B_0'A_0^{-1'} = A_{00}^{-1}B_{00}\Sigma_0 B_{00}'A_{00}^{-1'}$, then $\vartheta = \vartheta_0$.*

Proof: Let $\vartheta \neq \vartheta_0$. Assumption **A5** implies that either $A_0^{-1}B_0\Sigma B_0'A_0^{-1'} \neq A_{00}^{-1}B_{00}\Sigma_0 B_{00}'A_{00}^{-1'}$ or there exist matrices C_i such that $C_{i_0} \neq 0$ for some

Table 1: Empirical size of standard and modified tests: relative frequencies (in %) of rejection of $H_0 : b_1(2,2) = 0$. The number of replications is $N = 1000$.

Model	Length n	Level	Standard Test			Modified Test		
			Wald	LM	LR	Wald	LM	LR
		$\alpha = 1\%$	1.1	0.7	0.8	1.7	0.7	1.7
I	$n = 500$	$\alpha = 5\%$	5.0	4.5	5.1	6.0	5.2	6.0
		$\alpha = 10\%$	8.9	9.3	9.4	11.0	9.9	10.9
		$\alpha = 1\%$	0.7	0.8	0.7	1.0	0.6	1.0
I	$n = 2,000$	$\alpha = 5\%$	5.0	4.3	4.6	5.5	5.1	5.5
		$\alpha = 10\%$	9.2	8.6	8.8	10.0	9.0	10.2
		$\alpha = 1\%$	**0.0**	**0.0**	**0.0**	1.4	1.4	1.3
II	$n = 500$	$\alpha = 5\%$	**0.6**	**0.5**	**0.6**	6.2	**6.5**	6.1
		$\alpha = 10\%$	**2.3**	**2.2**	**2.2**	**12.0**	11.2	**12.0**
		$\alpha = 1\%$	**0.0**	**0.0**	**0.0**	0.9	0.7	0.9
II	$n = 2,000$	$\alpha = 5\%$	**0.4**	**0.3**	**0.3**	4.6	4.3	4.6
		$\alpha = 10\%$	**1.3**	**1.3**	**1.3**	9.2	9.8	9.2

I: Strong VARMA(1,1) model (9)-(10) with $\vartheta_0 = (0.95, 2, 0)$
II: Weak VARMA(1,1) model (9)-(11) with $\vartheta_0 = (0.95, 2, 0)$

$i_0 > 0$ and

$$A_0^{-1}B_0B_\vartheta^{-1}(z)A_\vartheta(z) - A_{00}^{-1}B_{00}B_{\vartheta_0}^{-1}(z)A_{\vartheta_0}(z) = \sum_{i=i_0}^{\infty} C_i z^i.$$

By contradiction, assume that $A_0^{-1}B_0B_\vartheta^{-1}(L)A_\vartheta(L)X_t = A_{00}^{-1}B_{00}\epsilon_t = A_{00}^{-1}B_{00}B_{\vartheta_0}^{-1}(L)A_{\vartheta_0}(L)X_t$ with probability one. This implies that there exists $\lambda \neq 0$ such that $\lambda' X_{t-i_0}$ is almost surely a linear combination of the components of $X_{t-i}, i > i_0$. By stationarity, it follows that $\lambda' X_t$ is almost surely a linear combination of the components of $X_{t-i}, i > 0$. Thus $\lambda'\epsilon_t = 0$ almost surely, which is impossible when the variance Σ_0 of ϵ_t is positive definite. □

Proof of Theorem 1: Note that, due to the initial conditions, $\{\tilde{e}_t(\vartheta)\}$ is not stationary, but can be approximated by the stationary ergodic process

$$e_t(\vartheta) \;=\; A_0^{-1}B_0B_\vartheta^{-1}(L)A_\vartheta(L)X_t. \tag{12}$$

Table 2: Empirical power of standard and modified tests: relative frequencies (in %) of rejection of $H_0 : b_1(2,2) = 0$. The number of replications is $N = 1000$.

Model	Length n	Level	Standard Test			Modified Test		
			Wald	LM	LR	Wald	LM	LR
		$\alpha = 1\%$	6.8	5.9	6.6	8.0	6.5	7.9
III	$n = 500$	$\alpha = 5\%$	20.5	19.4	20.4	21.6	20.1	21.7
		$\alpha = 10\%$	29.5	29.0	29.4	30.6	29.5	30.6
		$\alpha = 1\%$	1.7	1.8	1.7	15.5	14.3	15.6
IV	$n = 500$	$\alpha = 5\%$	11.4	9.4	10.1	35.1	34.0	35.0
		$\alpha = 10\%$	21.1	20.2	20.6	47.1	44.9	46.8

III: Strong VARMA(1,1) model (9)-(10) with $\vartheta_0 = (0.95, 2, 0.05)$
IV: Weak VARMA(1,1) model (9)-(11) with $\vartheta_0 = (0.95, 2, 0.05)$

From an extension of Lemma 1 in FZ, it is easy to show that $\sup_{\vartheta\in\Theta} \|\tilde{e}_t(\vartheta) - e_t(\vartheta)\| \to 0$ almost surely at an exponential rate, as $t \to \infty$. We thus have

$$\tilde{\ell}_n(\vartheta) \overset{o_P(1)}{=} \ell_n(\vartheta) := \frac{1}{n}\sum_{t=1}^{n} l_t(\vartheta) \quad \text{as} \quad n \to \infty,$$

where

$$l_t(\vartheta) = d\log(2\pi) + \log\det\Sigma_e + e_t'(\vartheta)\Sigma_e^{-1}e_t(\vartheta).$$

Now the ergodic theorem shows that almost surely

$$\ell_n(\vartheta) \to d\log(2\pi) + Q(\vartheta),$$

where $Q(\vartheta) = \log\det\Sigma_e + Ee_1'(\vartheta)\Sigma_e^{-1}e_1(\vartheta)$. We have

$$\begin{aligned} Q(\vartheta) &= E\left\{e_1(\vartheta_0)\right\}'\Sigma_e^{-1}\left\{e_1(\vartheta_0)\right\} + \log\det\Sigma_e \\ &\quad + E\left\{e_1(\vartheta) - e_1(\vartheta_0)\right\}'\Sigma_e^{-1}\left\{e_1(\vartheta) - e_1(\vartheta_0)\right\} \\ &\quad + 2E\left\{e_1(\vartheta) - e_1(\vartheta_0)\right\}'\Sigma_e^{-1}e_1(\vartheta_0). \end{aligned}$$

The last expectation is null because the linear innovation $e_t = e_t(\vartheta_0)$ is orthogonal to the linear past (*i.e.* to the Hilbert space H_{t-1} generated by linear

combinations of the X_u for $u < t$), and because $\{e_t(\vartheta) - e_t(\vartheta_0)\}$ belongs to this linear past H_{t-1}. Moreover

$$\begin{aligned} Q(\vartheta_0) &= \log\det\Sigma_{e0} + Ee_1'(\vartheta_0)\Sigma_{e0}^{-1}e_1(\vartheta_0) \\ &= \log\det\Sigma_{e0} + \mathrm{Tr}\,\Sigma_{e0}^{-1}Ee_1(\vartheta_0)e_1'(\vartheta_0) = \log\det\Sigma_{e0} + d. \end{aligned}$$

Thus

$$Q(\vartheta) - Q(\vartheta_0) \geq \mathrm{Tr}\,\Sigma_e^{-1}\Sigma_{e0} - \log\det\Sigma_e^{-1}\Sigma_{e0} - d \geq 0 \tag{13}$$

using the elementary inequality $\mathrm{Tr}(A^{-1}B) - \log\det(A^{-1}B) \geq \mathrm{Tr}(A^{-1}A) - \log\det(A^{-1}A) = d$ for all symmetric positive semi-definite matrices of order $d \times d$. At least one of the two inequalities in (13) is strict, unless if $e_1(\vartheta) = e_1(\vartheta_0)$ with probability 1 and $\Sigma_e = \Sigma_{e0}$, which is equivalent to $\vartheta = \vartheta_0$ by Lemma 1. The rest of the proof relies on standard compactness arguments, as in Theorem 1 of FZ. □

Proof of Theorem 2: In view of Theorem 1 and **A6**, we have almost surely $\hat{\vartheta}_n \to \vartheta_0 \in \overset{\circ}{\Theta}$. Thus $\partial\tilde{\ell}_n(\hat{\vartheta}_n)/\partial\vartheta = 0$ for sufficiently large n, and a Taylor expansion gives

$$0 \overset{o_P(1)}{=} \sqrt{n}\frac{\partial\ell_n(\vartheta_0)}{\partial\vartheta} + \frac{\partial^2\ell_n(\vartheta_0)}{\partial\vartheta\partial\vartheta'}\sqrt{n}\left(\hat{\vartheta}_n - \vartheta_0\right), \tag{14}$$

using arguments given in FZ (proof of Theorem 2). The proof then directly follows from Lemma 3 and Lemma 5 below. □

We first state elementary derivative rules, which can be found in Appendix A.13 of Lütkepohl (1993).

Lemma 2. *If $f(A)$ is a scalar function of a matrix A whose elements a_{ij} are function of a variable x, then*

$$\frac{\partial f(A)}{\partial x} = \sum_{i,j}\frac{\partial f(A)}{\partial a_{ij}}\frac{\partial a_{ij}}{\partial x} = Tr\left\{\frac{\partial f(A)}{\partial A'}\frac{\partial A}{\partial x}\right\}. \tag{15}$$

When A is invertible, we also have

$$\frac{\partial\log|\det(A)|}{\partial A'} = A^{-1} \tag{16}$$

$$\frac{\partial Tr(CA^{-1}B)}{\partial A'} = -A^{-1}BCA^{-1} \tag{17}$$

$$\frac{\partial Tr(CAB)}{\partial A'} = BC \tag{18}$$

Lemma 3. *Under the assumptions of Theorem 2, almost surely*

$$\frac{\partial^2 \ell_n(\vartheta_0)}{\partial\vartheta\partial\vartheta'} \to J,$$

where J is invertible.

Proof of Lemma 3: Let $\vartheta = (\vartheta_1, \ldots, \vartheta_{k_0})'$. In view of (15), (16) and (17), for all $i \in \{1, \ldots, k_0\}$, we have

$$\frac{\partial l_t(\vartheta)}{\partial\vartheta_i} = \mathrm{Tr}\left\{\Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_i} - \Sigma_e^{-1}e_t(\vartheta)e_t'(\vartheta)\Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_i}\right\} + 2\frac{\partial e_t'(\vartheta)}{\partial\vartheta_i}\Sigma_e^{-1}e_t(\vartheta). \quad (19)$$

Using the previous relations and (18), for all $i, j \in \{1, \ldots, k_0\}$, we have

$$\begin{aligned}\frac{\partial^2 l_t(\vartheta)}{\partial\vartheta_i\partial\vartheta_j} &= \mathrm{Tr}\left\{\Sigma_e^{-1}\frac{\partial^2\Sigma_e}{\partial\vartheta_i\partial\vartheta_j} - \Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_i}\Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_j} - \Sigma_e^{-1}e_t(\vartheta)e_t'(\vartheta)\Sigma_e^{-1}\frac{\partial^2\Sigma_e}{\partial\vartheta_i\partial\vartheta_j}\right.\\
&\quad +\Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_i}\Sigma_e^{-1}e_t(\vartheta)e_t'(\vartheta)\Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_j} + \Sigma_e^{-1}e_t(\vartheta)e_t'(\vartheta)\Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_i}\Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_j}\\
&\quad \left.-\Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_i}\Sigma_e^{-1}\frac{\partial e_t(\vartheta)e_t'(\vartheta)}{\partial\vartheta_j}\right\} + 2\frac{\partial^2 e_t'(\vartheta)}{\partial\vartheta_i\partial\vartheta_j}\Sigma_e^{-1}e_t(\vartheta)\\
&\quad +2\frac{\partial e_t'(\vartheta)}{\partial\vartheta_i}\Sigma_e^{-1}\frac{\partial e_t(\vartheta)}{\partial\vartheta_j} - 2\mathrm{Tr}\left\{\Sigma_e^{-1}e_t(\vartheta)\frac{\partial e_t'(\vartheta)}{\partial\vartheta_i}\Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_j}\right\}. \quad (20)\end{aligned}$$

Using $Ee_te_t' = \Sigma_e$, $Ee_t = 0$, the uncorrelatedness between e_t and the linear past H_{t-1}, $\partial e_t(\vartheta_0)/\partial\vartheta_i \in H_{t-1}$, and $\partial^2 e_t(\vartheta_0)/\partial\vartheta_i\partial\vartheta_j \in H_{t-1}$, we have

$$\begin{aligned}E\frac{\partial^2 l_t(\vartheta_0)}{\partial\vartheta_i\partial\vartheta_j} &= \mathrm{Tr}\left\{\Sigma_{e0}^{-1}\frac{\partial\Sigma_e(\vartheta_0)}{\partial\vartheta_i}\Sigma_{e0}^{-1}\frac{\partial\Sigma_e(\vartheta_0)}{\partial\vartheta_j}\right\} + 2E\frac{\partial e_t'(\vartheta_0)}{\partial\vartheta_i}\Sigma_{e0}^{-1}\frac{\partial e_t(\vartheta_0)}{\partial\vartheta_j}\\
&= J(i,j). \quad (21)\end{aligned}$$

The ergodic theorem and the next lemma conclude. □

Lemma 4. *Under the assumptions of Theorem 2, the matrix*

$$J = E\frac{\partial^2 l_t(\vartheta_0)}{\partial\vartheta\partial\vartheta'}$$

is invertible.

Proof of Lemma 4: In view of (21), we have $J = J_1 + J_2$, where

$$J_2 = 2E\frac{\partial e_t'(\vartheta_0)}{\partial \vartheta}\Sigma_{e0}^{-1}\frac{\partial e_t(\vartheta_0)}{\partial \vartheta'}$$

and

$$J_1(i,j) = \mathrm{Tr}\left\{\Sigma_{e0}^{-1/2}\frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_i}\Sigma_{e0}^{-1/2}\Sigma_{e0}^{-1/2}\frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_j}\Sigma_{e0}^{-1/2}\right\} = \mathbf{h}_i'\mathbf{h}_j,$$

with

$$\mathbf{h}_i = (\Sigma_{e0}^{-1/2}\otimes\Sigma_{e0}^{-1/2})\mathbf{d}_i, \quad \mathbf{d}_i = \mathrm{vec}\frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_i}.$$

In the previous derivations, we used the well-known relations $\mathrm{Tr}(A'B) = (\mathrm{vec}A)'\mathrm{vec}B$ and $\mathrm{vec}(ABC) = (C'\otimes A)\mathrm{vec}B$. Note that the matrices J, J_1 and J_2 are semi-definite positive. If J is singular, then there exists a vector $\boldsymbol{c} = (c_1,\ldots,c_{k_0})' \neq 0$ such that $\boldsymbol{c}'J_1\boldsymbol{c} = \boldsymbol{c}'J_2\boldsymbol{c} = 0$. Since $\Sigma_{e0}^{-1/2}\otimes\Sigma_{e0}^{-1/2}$ and Σ_{e0}^{-1} are definite positive, we have $\boldsymbol{c}'J_1\boldsymbol{c} = 0$ if and only if

$$\sum_{k=1}^{k_0} c_k\mathbf{d}_k = \sum_{k=1}^{k_0} c_k\mathrm{vec}\frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_k} = 0 \tag{22}$$

and $\boldsymbol{c}'J_2\boldsymbol{c} = 0$ if and only if $\sum_{k=1}^{k_0} c_k\frac{\partial e_t(\vartheta_0)}{\partial \vartheta_k} = 0$ *a.s.* Differentiating the two sides of the reduced form representation (3), the latter equation yields the VARMA$(p-1,q-1)$ equation $\sum_{i=1}^{p} A_i^* X_{t-i} = \sum_{j=1}^{p} B_j^* e_{t-j}$. The identifiability assumption **A5** excludes the existence of such a representation. Thus

$$A_i^* = \sum_{k=1}^{k_0} c_k\frac{\partial A_0^{-1}A_i}{\partial \vartheta_k}(\vartheta_0) = 0, \quad B_j^* = \sum_{k=1}^{k_0} c_k\frac{\partial A_0^{-1}B_jB_0^{-1}A_0}{\partial \vartheta_k}(\vartheta_0) = 0. \tag{23}$$

It can be seen that (22) and (23), for $i = 1,\ldots,p$ and $j = 1,\ldots,q$, are equivalent to $\dot{M}_{\vartheta_0}\,\boldsymbol{c} = 0$. We conclude from **A8**. □

Lemma 5. *Under the assumptions of Theorem 2,*

$$\sqrt{n}\frac{\partial \ell_n(\vartheta_0)}{\partial \vartheta} \xrightarrow{\mathcal{L}} \mathcal{N}(0, I).$$

Proof of Lemma 5: In view of (12), we have

$$\frac{\partial e_t(\vartheta_0)}{\partial \vartheta_i} = \sum_{\ell=1}^{\infty} d'_\ell e_{t-\ell}, \tag{24}$$

where the sequence of matrices $d_\ell = d_\ell(i)$ is such that $\|d_\ell\| \to 0$ at a geometric rate as $\ell \to \infty$. By (19), we have for all m

$$\begin{aligned} \frac{\partial l_t(\vartheta_0)}{\partial \vartheta_i} &= \mathrm{Tr}\left[\Sigma_{e0}^{-1}\left\{I_d - e_t e_t' \Sigma_{e0}^{-1}\right\}\frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_i}\right] + 2\frac{\partial e_t'(\vartheta_0)}{\partial \vartheta_i}\Sigma_{e0}^{-1} e_t \\ &= Y_{t,m,i} + Z_{t,m,i} \end{aligned}$$

where

$$\begin{aligned} Y_{t,m,i} &= \mathrm{Tr}\left[\Sigma_{e0}^{-1}\left\{I_d - e_t e_t' \Sigma_{e0}^{-1}\right\}\frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_i}\right] + 2\sum_{\ell=1}^{m} e'_{t-\ell} d'_\ell \Sigma_{e0}^{-1} e_t \\ Z_{t,m,i} &= 2\sum_{\ell=m+1}^{\infty} e'_{t-\ell} d'_\ell \Sigma_{e0}^{-1} e_t. \end{aligned}$$

Let $Y_{t,m} = (Y_{t,m,1}, \ldots, Y_{t,m,k_0})'$ and $Z_{t,m} = (Z_{t,m,1}, \ldots, Y_{t,m,k_0})'$. The processes $(Y_{t,m})_t$ and $(Z_{t,m})_t$ are stationary and centered. Moreover, under Assumption **A7** and m fixed, the process $Y = (Y_{t,m})_t$ is strongly mixing, with mixing coefficients $\alpha_Y(h) \leq \alpha_\epsilon(\max\{0, h-m\})$. Applying the central limit theorem (CLT) for mixing processes (see Herrndorf, 1984) we directly obtain

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Y_{t,m} \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_m), \quad I_m = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\left(Y_{t,m}, Y_{t-h,m}\right).$$

As in FZ Lemma 3, one can show that $I = \lim_{m\to\infty} I_m$ exists. Since $\|Z_{t,m}\|_2 \to 0$ at an exponential rate when $m \to \infty$, using the arguments given in FZ Lemma 4, one can show that

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left\{\left\| n^{-1/2}\sum_{t=1}^{n} Z_{t,m}\right\| > \varepsilon\right\} = 0 \tag{25}$$

for every $\varepsilon > 0$. From a standard result (see *e.g.* Brockwell and Davis, 1991, Proposition 6.3.9), we deduce that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\vartheta_0)}{\partial \vartheta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} Y_{t,m} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{t,m} \xrightarrow{\mathcal{L}} \mathcal{N}(0, I),$$

which completes the proof. □

Proof of Theorem 3: Note that

$$\tilde{\ell}_n(\vartheta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\vartheta), \quad \tilde{l}_t(\vartheta) = d\log(2\pi) + \log\det\Sigma_e + \tilde{e}_t'(\vartheta)\Sigma_e^{-1}\tilde{e}_t(\vartheta).$$

Under the assumption of the theorem, $\partial\tilde{e}_t'(\vartheta)/\partial\vartheta^{(2)} = 0$, and (19) yields

$$\frac{\partial\tilde{l}_t(\hat{\vartheta}_n)}{\partial\vartheta_i} = \mathrm{Tr}\left[\hat{\Sigma}_e^{-1}\left\{I_d - \tilde{e}_t(\hat{\vartheta}_n^{(1)})\tilde{e}_t'(\hat{\vartheta}_n^{(1)})\hat{\Sigma}_e^{-1}\right\}\frac{\partial\Sigma_e(\hat{\vartheta}_n)}{\partial\vartheta_i}\right]$$

for $i = k_1 + 1, \ldots k_0$, with $\hat{\Sigma}_e$ such that $\hat{\vartheta}_n^{(2)} = D\,\mathrm{vec}\,\hat{\Sigma}_e$. Assumption **A6** entails that the first order condition $\partial\tilde{\ell}_n(\hat{\vartheta}_n)/\partial\vartheta^{(2)} = 0$ is satisfied for n large enough. We then have

$$\hat{\Sigma}_e = n^{-1}\sum_{t=1}^{n}\tilde{e}_t(\hat{\vartheta}_n^{(1)})\tilde{e}_t'(\hat{\vartheta}_n^{(1)})$$

and

$$\tilde{\ell}_n(\hat{\vartheta}_n) = d\log(2\pi) + \log\det\hat{\Sigma}_e + d,$$

because

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{e}_t'(\hat{\vartheta}_n^{(1)})\hat{\Sigma}_e^{-1}\tilde{e}_t(\hat{\vartheta}_n^{(1)}) = \mathrm{Tr}\left[\frac{1}{n}\sum_{t=1}^{n}\tilde{e}_t(\hat{\vartheta}_n^{(1)})\tilde{e}_t'(\hat{\vartheta}_n^{(1)})\hat{\Sigma}_e^{-1}\right] = d.$$

The conclusion follows. □

References

[1] Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

[2] Berk, K.N. (1974) Consistent Autoregressive Spectral Estimates. *Annals of Statistics*, **2**, 489–502.

[3] Brockwell, P.J. and Davis, R.A. (1991) *Time series: theory and methods.* Springer Verlag, New York.

[4] den Hann, W.J. and Levin, A. (1997) A Practitioner's Guide to Robust Covariance Matrix Estimation. *In Handbook of Statistics* 15, Rao, C.R. and G.S. Maddala (eds), 291–341.

[5] Duchesne, P. and Roy, R. (2004) On consistent testing for serial correlation of unknown form in vector time series models. *Journal of Multivariate Analysis* 89, 148–180.

[6] Dufour, J-M., and Pelletier, D. (2005) Practical methods for modelling weak VARMA processes: identification, estimation and specification with a macroeconomic application. *Technical report, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada.*

[7] Dunsmuir, W.T.M. and Hannan, E.J. (1976) Vector linear time series models, *Advances in Applied Probability* 8, 339–364.

[8] Fan, J. and Yao, Q. (2003) *Nonlinear time series: Nonparametric and parametric methods.* Springer Verlag, New York.

[9] Francq, C. and Raïssi, H. (2007) Multivariate Portmanteau Test for Autoregressive Models with Uncorrelated but Nonindependent Errors, *Journal of Time Series Analysis* 28, 454–470.

[10] Francq, C., Roy, R. and Zakoïan, J-M. (2005) Diagnostic checking in ARMA Models with Uncorrelated Errors, *Journal of the American Statistical Association* 100, 532–544.

[11] Francq, C. and Zakoïan, J-M. (1998) Estimating linear representations of nonlinear processes, *Journal of Statistical Planning and Inference* 68, 145–165.

[12] Francq, and Zakoïan, J-M. (2005) Recent results for linear time series models with non independent innovations. In *Statistical Modeling and Analysis for Complex Data Problems,* Chap. 12 (eds P. DUCHESNE and B. RÉMILLARD). New York: Springer Verlag, 137–161.

[13] Francq, and Zakoïan, J-M. (2007) HAC estimation and strong linearity testing in weak ARMA models, *Journal of Multivariate Analysis* 98, 114–144.

[14] Hannan, E.J. (1976) The identification and parametrization of ARMAX and state space forms, *Econometrica* 44, 713–723.

[15] Hannan, E.J. and Deistler, M. (1988) *The Statistical Theory of Linear Systems* John Wiley, New York.

[16] Hannan, E.J., Dunsmuir, W.T.M. and Deistler, M. (1980) Estimation of vector ARMAX models, *Journal of Multivariate Analysis* 10, 275–295.

[17] Hannan, E.J. and Rissanen (1982) Recursive estimation of mixed of Autoregressive Moving Average order, *Biometrika* 69, 81–94.

[18] Herrndorf, N. (1984) A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability* 12, 141–153.

[19] Imhof, J.P. (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419–426.

[20] Kascha, C. (2007) A Comparison of Estimation Methods for Vector Autoregressive Moving-Average Models. *ECO Working Papers, EUI ECO 2007/12,* http://hdl.handle.net/1814/6921.

[21] Lütkepohl, H. (1993) *Introduction to multiple time series analysis.* Springer Verlag, Berlin.

[22] Lütkepohl, H. (2005) *New introduction to multiple time series analysis.* Springer Verlag, Berlin.

[23] Newey, W.K., and West, K.D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

[24] Nsiri, S., and Roy, R.] (1996) Identification of refined ARMA echelon form models for multivariate time series. *Journal of Multivariate Analysis* 56, 207–231.

[25] Reinsel, G.C. (1997) *Elements of multivariate time series Analysis.* Second edition. Springer Verlag, New York.

[26] Rissanen, J. and Caines, P.E. (1979) The strong consistency of maximum likelihood estimators for ARMA processes. *Annals of Statistics* 7, 297–315.

[27] Romano, J.L. and Thombs, L.A. (1996) Inference for autocorrelations under weak assumptions, *Journal of the American Statistical Association* 91, 590–600.

[28] Tong, H. (1990) Non-linear time series: *A Dynamical System Approach,* Clarendon Press Oxford.

[30] van der Vaart, A.W. (1998) Asymptotic statistics, *Cambridge University Press,* Cambridge.

[30] van der Vaart, A.W. (1998) Asymptotic statistics, *Cambridge University Press,* Cambridge.

[31] Wald, A. (1949) Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* 20, 595–601.

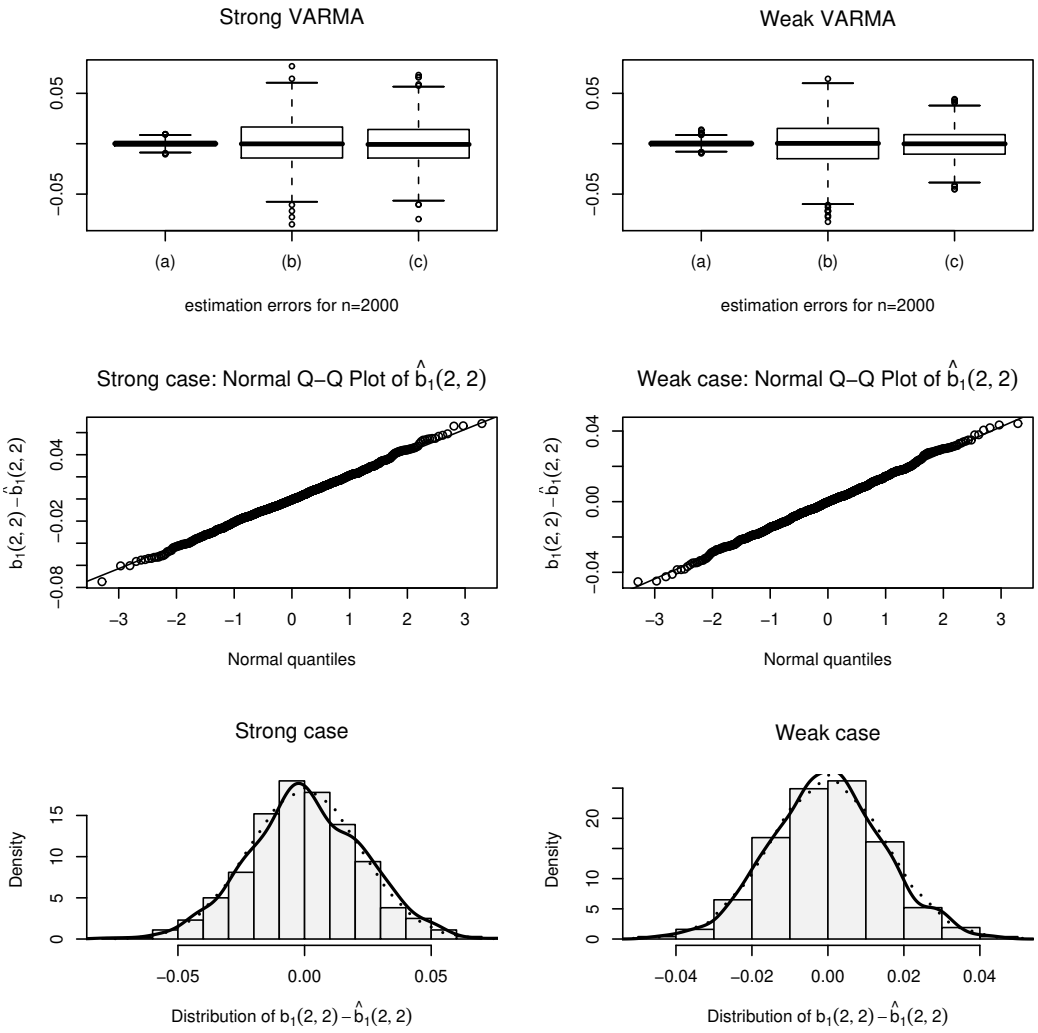

Figure 1: QMLE of $N = 1,000$ independent simulations of the VARMA(1,1) model (9) with size $n = 2,000$ and unknown parameter $\vartheta_0 = ((a_1(2,2), b_1(2,1), b_1(2,2)) = (0.95, 2, 0)$, when the noise is strong (left panels) and when the noise is the weak noise (11) (right panels). Points (a)-(c), in the box-plots of the top panels, display the distribution of the estimation errors $\hat{\vartheta}(i) - \vartheta_0(i)$ for $i = 1, 2, 3$. The panels of the middle present the Q-Q plot of the estimates $\hat{\vartheta}(3) = \hat{b}_1(2,2)$ of the last parameter. The bottom panels display the distribution of the same estimates. The kernel density estimate is displayed in full line, and the centered Gaussian density with the same variance is plotted in dotted line.

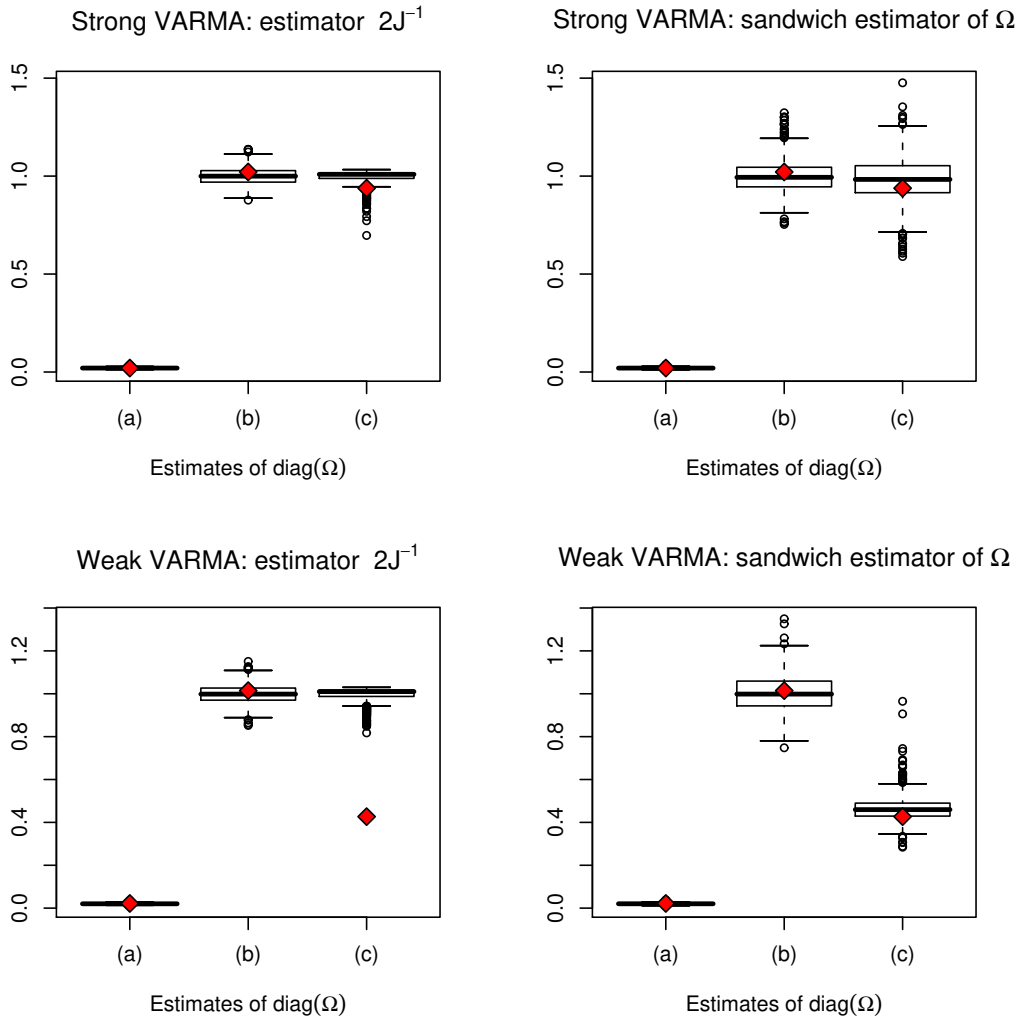

Figure 2: Comparison of standard and modified estimates of the asymptotic variance Ω of the QMLE, on the simulated models presented in Figure 1. The diamond symbols represent the mean, over the $N = 1,000$ replications, of the standardized squared errors $n\{\hat{a}_1(2,2) - 0.95\}^2$ for (a) (0.02 in the strong and weak cases), $n\left\{\hat{b}_1(2,1) - 2\right\}^2$ for (b) (1.02 in the strong case and 1.01 in the strong case) and $n\left\{\hat{b}_1(2,2)\right\}^2$ for (c) (0.94 in the strong case and 0.43 in the weak case).

Estimating structural VARMA models with uncorrelated but non-independent error terms: **Complementary results that are not submitted for publication**

A. Additional example

Example 2. Denoting by $a_{0i}(k,\ell)$ and $b_{0i}(k,\ell)$ the generic elements of the matrices A_{0i} and B_{0i}, the Kronecker indices are defined by $p_k = \max\{i : a_{0i}(k,\ell) \neq 0 \text{ or } b_{0i}(k,\ell) \neq 0 \text{ for some } \ell = 1,\dots,d\}$. To ensure relatively parsimonious parameterizations, one can specify an echelon form depending on the Kronecker indices $(p_1,\dots,p_d)$. The reader is refereed to Lütkepohl (1993) for details about the echelon form. For instance, a 3-variate ARMA process with Kronecker indices $(1,2,0)$ admits the echelon form

$$\begin{aligned} &\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \times & \times & 1 \end{pmatrix} X_t - \begin{pmatrix} \times & \times & 0 \\ 0 & \times & 0 \\ 0 & 0 & 0 \end{pmatrix} X_{t-1} - \begin{pmatrix} 0 & 0 & 0 \\ \times & \times & 0 \\ 0 & 0 & 0 \end{pmatrix} X_{t-2} \\ = &\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \times & \times & 1 \end{pmatrix} \epsilon_t - \begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & 0 & 0 \end{pmatrix} \epsilon_{t-1} - \begin{pmatrix} 0 & 0 & 0 \\ \times & \times & \times \\ 0 & 0 & 0 \end{pmatrix} \epsilon_{t-2} \end{aligned}$$

where $\times$ denotes an unconstrained element. The variance of ϵ_t is defined by 6 additional parameters. This echelon form thus corresponds to a parametrization by a vector ϑ of size $k_0 = 24$.

B. Verification of Assumption A8 on Example 1

In this example, we have

$$M_{\vartheta_0} = \begin{pmatrix} \alpha_{01} & \alpha_{02} & \sigma_{01}^2 & \sigma_{01}^2\alpha_{03} \\ \alpha_{01}\alpha_{03}+\alpha_{04} & \alpha_{02}\alpha_{03}+\alpha_{05} & \sigma_{01}^2\alpha_{03} & \sigma_{01}^2\alpha_{03}^2+\sigma_{02}^2 \end{pmatrix}.$$

Thus

$$\dot{M}_{\vartheta_0} = \begin{pmatrix} 1 & \alpha_{03} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \alpha_{03} & 0 & 0 & 0 & 0 \\ 0 & \alpha_{01} & 0 & \alpha_{02} & 0 & \sigma_{01}^2 & \sigma_{01}^2 & 2\alpha_{03}\sigma_{01}^2 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \alpha_{03} & \alpha_{03} & \alpha_{03}^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

is of full rank $k_0 = 7$.

C. Details on the proof of Theorem 1

Lemma 6. *Under the assumptions of Theorem 1, we have*

$$\sup_{\vartheta\in\Theta} \|\tilde{e}_t(\vartheta) - e_t(\vartheta)\| \leq K\rho^t,$$

where ρ is a constant belonging to $[0,1)$, and $K > 0$ is measurable with respect to the σ-field generated by $\{X_u, u \leq 0\}$.

Proof of Lemma 6: We have

$$e_t(\vartheta) = X_t - \sum_{i=1}^{p} A_0^{-1}A_iX_{t-i} + \sum_{i=1}^{q} A_0^{-1}B_iB_0^{-1}A_0e_{t-i}(\vartheta) \qquad \forall t \in \mathbb{Z}, \quad (26)$$

and

$$\tilde{e}_t(\vartheta) = X_t - \sum_{i=1}^{p} A_0^{-1}A_iX_{t-i} + \sum_{i=1}^{q} A_0^{-1}B_iB_0^{-1}A_0\tilde{e}_{t-i}(\vartheta) \qquad t = 1,\dots,n \quad (27)$$

with the initial values $\tilde{e}_0(\vartheta) = \cdots = \tilde{e}_{1-q}(\vartheta) = X_0 = \cdots = X_{1-p} = 0$. Let

$$\underline{e}_t(\vartheta) = \begin{pmatrix} e_t(\vartheta) \\ e_{t-1}(\vartheta) \\ \vdots \\ e_{t-q+1}(\vartheta) \end{pmatrix}, \qquad \underline{\tilde{e}}_t(\vartheta) = \begin{pmatrix} \tilde{e}_t(\vartheta) \\ \tilde{e}_{t-1}(\vartheta) \\ \vdots \\ \tilde{e}_{t-q+1}(\vartheta) \end{pmatrix}.$$

From (26) and (27), we have

$$\underline{e}_t(\vartheta) = \underline{b}_t + \boldsymbol{C}\underline{e}_{t-1}(\vartheta) \qquad \forall t \in \mathbb{Z},$$

and

$$\underline{\tilde{e}}_t(\vartheta) = \underline{\tilde{b}}_t + \boldsymbol{C}\underline{\tilde{e}}_{t-1}(\vartheta) \qquad t = 1,\dots,n,$$

where

$$\boldsymbol{C} = \begin{pmatrix} A_0^{-1}B_1B_0^{-1}A_0 & \cdots & A_0^{-1}B_qB_0^{-1}A_0 \\ I_{(q-1)d} & & 0_{(q-1)d} \end{pmatrix}, \quad \underline{b}_t = \begin{pmatrix} X_t - \sum_{i=1}^{p} A_0^{-1}A_iX_{t-i} \\ 0_d \\ \vdots \\ 0_d \end{pmatrix},$$

$\tilde{\underline{b}}_t = \underline{b}_t$ for $t > p$, $\tilde{b}_t = 0_{qd}$ for $t \leq 0$, and

$$\tilde{\underline{b}}_t = \begin{pmatrix} X_t - \sum_{i=1}^{t-1} A_0^{-1} A_i X_{t-i} \\ 0_d \\ \vdots \\ 0_d \end{pmatrix} \text{ for } t = 1, \ldots, p.$$

Writing $\underline{d}_t(\vartheta) = \underline{e}_t(\vartheta) - \tilde{\underline{e}}_t(\vartheta)$, we obtain for $t > p$,

$$\begin{aligned} \underline{d}_t(\vartheta) &= \boldsymbol{C}\underline{d}_{t-1}(\vartheta) = \boldsymbol{C}^{t-p}\underline{d}_p(\vartheta) \\ &= \boldsymbol{C}^{t-p}\left\{\left(\underline{b}_p - \tilde{\underline{b}}_p\right) + \boldsymbol{C}\left(\underline{b}_{p-1} - \tilde{\underline{b}}_{p-1}\right) + \cdots + \boldsymbol{C}^{p-1}\left(\underline{b}_1 - \tilde{\underline{b}}_1\right)\right\} + \boldsymbol{C}^p \underline{b}_0. \end{aligned}$$

Note that $\boldsymbol{C}$ is the companion matrix of the polynomial

$$\mathcal{P}(z) = I_d - \sum_{i=1}^{q} A_0^{-1} B_i B_0^{-1} A_0 z^i = A_0^{-1} B_\vartheta(z) B_0^{-1} A_0.$$

By **A2**, the zeroes of $\mathcal{P}(z)$ are of modulus strictly greater than one:

$$\mathcal{P}(z) = 0 \Rightarrow |z| > 1 \tag{28}$$

By a well-known result on companion matrices, (28) is equivalent to $\rho(\boldsymbol{C}) < 1$, where $\rho(\boldsymbol{C})$ denote the spectral radius of $\boldsymbol{C}$. By the compactness of Θ, we thus have

$$\sup_{\vartheta \in \Theta} \rho(\boldsymbol{C}) < 1.$$

We thus have

$$\sup_{\vartheta \in \Theta} \|\underline{d}_t(\vartheta)\| \leq K\rho^t,$$

where K and ρ are as in the statement of the lemma. The conclusion follows. □

Lemma 7. *Under the assumptions of Theorem 1, we have*

$$\sup_{\vartheta \in \Theta} \left|\tilde{\ell}_n(\vartheta) - \ell_n(\vartheta)\right| = o(1)$$

almost surely.

Proof of Lemma 7: We have

$$\tilde{\ell}_n(\vartheta) - \ell_n(\vartheta) = \frac{1}{n}\sum_{t=1}^{n} \left\{\tilde{e}_t(\vartheta) - e_t(\vartheta)\right\}' \Sigma_e^{-1} \tilde{e}_t(\vartheta) + e_t'(\vartheta)\Sigma_e^{-1}\left\{\tilde{e}_t(\vartheta) - e_t(\vartheta)\right\}.$$

In the proof of this lemma and in the rest on the paper, the letters K and ρ denote generic constants, whose values can be modified along the text, such that $K > 0$ and $0 < \rho < 1$. By Lemma 6,

$$\begin{aligned} \sup_{\vartheta\in\Theta}\left|\tilde{\ell}_n(\vartheta) - \ell_n(\vartheta)\right| &\leq \frac{K}{n}\sum_{t=1}^{n}\rho^t\left(\sup_{\vartheta\in\Theta}\|e_t(\vartheta)\| + \sup_{\vartheta\in\Theta}\|\tilde{e}_t(\vartheta)\|\right) \\ &\leq \frac{K}{n}\sum_{t=1}^{n}\rho^t \sup_{\vartheta\in\Theta}\|e_t(\vartheta)\| \end{aligned} \tag{29}$$

In view of (26), and using **A1** and the compactness of Θ, we have

$$e_t(\vartheta) = X_t + \sum_{i=1}^{\infty} C_i(\vartheta)X_{t-i}, \qquad \sup_{\vartheta\in\Theta}\|C_i(\vartheta)\| \leq K\rho^i. \tag{30}$$

We thus have $E\sup_{\vartheta\in\Theta}\|e_t(\vartheta)\| < \infty$, and the Markov inequality entails

$$\sum_{t=1}^{\infty} P\left(\rho^t \sup_{\vartheta\in\Theta}\|e_t(\vartheta)\| > \varepsilon\right) \leq E\sup_{\vartheta\in\Theta}\|e_t(\vartheta)\| \sum_{t=1}^{\infty}\frac{\rho^t}{\varepsilon} < \infty.$$

By the Borel-Cantelli theorem, $\rho^t \sup_{\vartheta\in\Theta}\|e_t(\vartheta)\| \to 0$ almost surely as $t \to \infty$. The Cesàro theorem implies that the right-hand side of (29) converges to zero almost surely. □

Lemma 8. *Under the assumptions of Theorem 1, any $\vartheta \neq \vartheta_0$ has a neighborhood $V(\vartheta)$ such that*

$$\liminf_{n\to\infty} \inf_{\vartheta^*\in V(\vartheta)} \tilde{\ell}_n(\vartheta^*) > El_1(\vartheta_0), \quad a.s. \tag{31}$$

Moreover for any neighborhood $V(\vartheta_0)$ of ϑ_0 we have

$$\limsup_{n\to\infty} \inf_{\vartheta^*\in V(\vartheta_0)} \tilde{\ell}_n(\vartheta^*) \leq El_1(\vartheta_0), \quad a.s. \tag{32}$$

Proof of Lemma 8: For any $\vartheta \in \Theta$ and any positive integer k, let $V_k(\vartheta)$ be the open ball with center ϑ and radius $1/k$. Using Lemma 7, we have

$$\begin{aligned}
\liminf_{n\to\infty} \inf_{\vartheta^* \in V_k(\vartheta)\cap\Theta} \tilde{\ell}_n(\vartheta^*) &\geq \liminf_{n\to\infty} \inf_{\vartheta^* \in V_k(\vartheta)\cap\Theta} \ell_n(\vartheta^*) - \limsup_{n\to\infty} \sup_{\vartheta\in\Theta} |\ell_n(\vartheta) - \tilde{\ell}_n(\vartheta)| \\
&\geq \liminf_{n\to\infty} n^{-1} \sum_{t=1}^{n} \inf_{\vartheta^* \in V_k(\vartheta)\cap\Theta} l_t(\vartheta^*) \\
&= E \inf_{\vartheta^* \in V_k(\vartheta)\cap\Theta} l_1(\vartheta^*)
\end{aligned}$$

For the last equality we applied the ergodic theorem to the ergodic stationary process $\left\{\inf_{\vartheta^* \in V_k(\vartheta)\cap\Theta} \ell_t(\vartheta^*)\right\}_t$. By the Beppo-Levi theorem, when k increases to ∞, $E \inf_{\vartheta^* \in V_k(\vartheta)\cap\Theta} l_1(\vartheta^*)$ increases to $El_1(\vartheta)$. Because $El_1(\vartheta) = d\log(2\pi) + Q(\vartheta)$, the discussion which follows (13) entails $El_1(\vartheta) > El_1(\vartheta_0)$, and (31) follows.

To show (32), it suffices to remark that Lemma 7 and the ergodic theorem entail

$$\begin{aligned}
\limsup_{n\to\infty} \inf_{\vartheta^* \in V(\vartheta)\cap\Theta} \tilde{\ell}_n(\vartheta^*) &\leq \limsup_{n\to\infty} \inf_{\vartheta^* \in V(\vartheta)\cap\Theta} \ell_n(\vartheta^*) + \limsup_{n\to\infty} \sup_{\vartheta\in\Theta} |\ell_n(\vartheta) - \tilde{\ell}_n(\vartheta)| \\
&\leq \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} l_t(\vartheta_0) \\
&= El_1(\vartheta_0).
\end{aligned}$$

□

The proof of Theorem 1 is completed by the arguments of Wald (1949). More precisely, the compact set Θ is covered by a neighborhood $V(\vartheta_0)$ of ϑ_0 and a finite number of neighborhoods $V(\vartheta_1), \ldots, V(\vartheta_k)$ satisfying (31) with ϑ replaced by ϑ_i, $i = 1, \ldots, k$. In view of (31) and (32), we have almost surely

$$\inf_{\vartheta\in\Theta} \tilde{\ell}_n(\vartheta) = \min_{i=0,1,\ldots,k} \inf_{\vartheta\in V(\vartheta_i)\cap\Theta} \tilde{\ell}_n(\vartheta) = \inf_{\vartheta\in V(\vartheta_0)\cap\Theta} \tilde{\ell}_n(\vartheta)$$

for n large enough. Since the neighborhood $V(\vartheta_0)$ can be chosen arbitrarily small, the conclusion follows.

D. Details on the proof of Theorem 2

Lemma 9. *Under the assumptions of Theorem 2, we have*

$$\sqrt{n}\sup_{\vartheta\in\Theta}\left\|\frac{\partial\tilde{\ell}_n(\vartheta)}{\partial\vartheta}-\frac{\partial\ell_n(\vartheta)}{\partial\vartheta}\right\|=o_P(1).$$

Proof of Lemma 9: Similar to (30), Assumption **A1** entails that, for $k=1,\ldots,k_0$,

$$\frac{\partial e_t(\vartheta)}{\partial\vartheta_k}=\sum_{i=1}^{\infty}C_i^{(k)}(\vartheta)X_{t-i},\qquad \frac{\partial\tilde{e}_t(\vartheta)}{\partial\vartheta_k}=\sum_{i=1}^{t-1}C_i^{(k)}(\vartheta)X_{t-i}, \tag{33}$$

$$C_i^{(k)}(\vartheta)=\frac{\partial C_i(\vartheta)}{\partial\vartheta_k},\qquad \sup_{\vartheta\in\Theta}\left\|C_i^{(k)}(\vartheta)\right\|\leq K\rho^i. \tag{34}$$

Similar to Lemma 6, we then have

$$\sup_{\vartheta\in\Theta}\left\|\frac{\partial\tilde{e}_t(\vartheta)}{\partial\vartheta'}-\frac{\partial e_t(\vartheta)}{\partial\vartheta'}\right\|\leq K\rho^t. \tag{35}$$

Using (19), we have

$$\sqrt{n}\left\{\frac{\partial\tilde{\ell}_n(\vartheta)}{\partial\vartheta_k}-\frac{\partial\ell_n(\vartheta)}{\partial\vartheta_k}\right\}=a_1+a_2,$$

with

$$\begin{aligned}
a_1 &= \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\left\{\left(\frac{\partial e_t'(\vartheta)}{\partial\vartheta_k}-\frac{\partial\tilde{e}_t'(\vartheta)}{\partial\vartheta_k}\right)\Sigma_e^{-1}e_t(\vartheta)+\frac{\partial\tilde{e}_t'(\vartheta)}{\partial\vartheta_k}\Sigma_e^{-1}\left(e_t(\vartheta)-\tilde{e}_t(\vartheta)\right)\right\}\\
a_2 &= \mathrm{Tr}\left(\Sigma_e^{-1}\left[\left\{\tilde{e}_t(\vartheta)-e_t(\vartheta)\right\}\tilde{e}_t'(\vartheta)+e_t(\vartheta)\left\{\tilde{e}_t(\vartheta)-e_t(\vartheta)\right\}'\right]\Sigma_e^{-1}\frac{\partial\Sigma_e}{\partial\vartheta_k}\right).
\end{aligned}$$

From Lemma 6 and (35), it follows that

$$|a_1|+|a_2|\leq\frac{K}{\sqrt{n}}\sum_{t=1}^{n}b_t,\qquad b_t=\rho^t\left\{\sup_{\vartheta\in\Theta}\left\|\frac{\partial e_t(\vartheta)}{\partial\vartheta_k}\right\|+\sup_{\vartheta\in\Theta}\|e_t(\vartheta)\|+K\rho^t\right\}.$$

In view of (30) and (33)-(34), we have

$$E\sup_{\vartheta\in\Theta}\left\|\frac{\partial e_t(\vartheta)}{\partial\vartheta_k}\right\|<\infty,\quad E\sup_{\vartheta\in\Theta}\|e_t(\vartheta)\|<\infty.$$

By the Markov inequality, for all $\varepsilon > 0$ we thus have

$$P\left(\frac{K}{\sqrt{n}}\sum_{t=1}^{n} b_t > \varepsilon\right) \leq \frac{K}{\varepsilon\sqrt{n}}\sum_{t=1}^{n}\rho^t \to 0,$$

and the conclusion follows. □

Lemma 10. *Under the assumptions of Theorem 2, (14) holds, that is*

$$0 \overset{o_P(1)}{=} \sqrt{n}\frac{\partial \ell_n(\vartheta_0)}{\partial \vartheta} + \frac{\partial^2 \ell_n(\vartheta_0)}{\partial \vartheta \partial \vartheta'}\sqrt{n}\left(\hat{\vartheta}_n - \vartheta_0\right).$$

Proof of Lemma 10: By Lemma 9 and already given arguments, we have

$$0 = \sqrt{n}\frac{\tilde{\ell}_n(\hat{\vartheta}_n)}{\partial \vartheta} = \sqrt{n}\frac{\ell_n(\hat{\vartheta}_n)}{\partial \vartheta} + \sqrt{n}\frac{\tilde{\ell}_n(\hat{\vartheta}_n)}{\partial \vartheta} - \sqrt{n}\frac{\ell_n(\hat{\vartheta}_n)}{\partial \vartheta} \overset{o_P(1)}{=} \sqrt{n}\frac{\ell_n(\hat{\vartheta}_n)}{\partial \vartheta}.$$

Taylor expansions of the functions $\partial \ell_n(\cdot)/\partial \vartheta_i$ around ϑ_0 give

$$\sqrt{n}\frac{\ell_n(\hat{\vartheta}_n)}{\partial \vartheta} = \sqrt{n}\frac{\partial \ell_n(\vartheta_0)}{\partial \vartheta} + \left[\frac{\partial^2 \ell_n(\vartheta_i^*)}{\partial \vartheta_i \partial \vartheta_j}\right]\sqrt{n}\left(\hat{\vartheta}_n - \vartheta_0\right),$$

for some ϑ_i^* between $\hat{\vartheta}_n$ and ϑ_0.

From (20) and Lemma 2, one can see that the random terms involved in

$$\frac{\partial^3 l_t(\vartheta)}{\partial \vartheta_{k_1}\partial \vartheta_{k_2}\partial \vartheta_{k_3}}$$

are linear combinations of the elements of the following matrices

$$e_t(\vartheta)e_t'(\vartheta), \qquad e_t(\vartheta)\frac{\partial e_t'(\vartheta)}{\partial \vartheta_{k_i}}, \qquad e_t(\vartheta)\frac{\partial^2 e_t'(\vartheta)}{\partial \vartheta_{k_i}\partial \vartheta_{k_j}},$$
$$\frac{\partial e_t(\vartheta)}{\partial \vartheta_{k_i}}\frac{\partial e_t'(\vartheta)}{\partial \vartheta_{k_j}}, \qquad \frac{\partial e_t(\vartheta)}{\partial \vartheta_{k_i}}\frac{\partial^2 e_t'(\vartheta)}{\partial \vartheta_{k_j}\partial \vartheta_{k_\ell}}, \qquad e_t(\vartheta)\frac{\partial^3 e_t'(\vartheta)}{\partial \vartheta_{k_i}\partial \vartheta_{k_j}\partial \vartheta_{k_\ell}}.$$

Since all these matrices have finite expectations uniformly in $\vartheta \in \Theta$, we have

$$E\sup_{\vartheta\in\Theta}\left|\frac{\partial^3 l_t(\vartheta)}{\partial \vartheta_{k_1}\partial \vartheta_{k_2}\partial \vartheta_{k_3}}\right| < \infty \tag{36}$$

for all $k_1, k_2, k_3 \in \{1, \ldots, k_0\}$. Now a second Taylor expansion yields

$$\begin{aligned}\frac{\partial^2 \ell_n(\vartheta_i^*)}{\partial\vartheta_i\partial\vartheta_j} &= \frac{\partial^2 \ell_n(\vartheta_0)}{\partial\vartheta_i\partial\vartheta_j} + \frac{\partial}{\partial\vartheta'}\left\{\frac{\partial^2 \ell_n(\vartheta^*)}{\partial\vartheta_i\partial\vartheta_j}\right\}(\vartheta_i^* - \vartheta_0) \\ &= \frac{\partial^2 \ell_n(\vartheta_0)}{\partial\vartheta_i\partial\vartheta_j} + o(1) \quad \text{a.s.}\end{aligned}$$

using the ergodic theorem, (36) and the almost sure convergence of ϑ_i^* to ϑ_0. The proof is complete. □

Now we need the following covariance inequality obtained by Davydov (1968).[2]

Lemma 11 (Davydov (1968)). *Let p, q and r three positive numbers such that $p^{-1} + q^{-1} + r^{-1} = 1$. Davydov (1968) showed that*

$$|Cov(X, Y)| \leq K_0 \|X\|_p \|Y\|_q \left[\alpha\left\{\sigma(X), \sigma(Y)\right\}\right]^{1/r}, \tag{37}$$

where $\|X\|_p^p = EX^p$, K_0 is an universal constant, and $\alpha\left\{\sigma(X), \sigma(Y)\right\}$ denotes the strong mixing coefficient between the σ-fields $\sigma(X)$ and $\sigma(Y)$ generated by the random variables X and Y, respectively.

Lemma 12. *Under the assumptions of Theorem 2, (25) holds, that is*

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left\{\left\| n^{-1/2} \sum_{t=1}^{n} Z_{t,m} \right\| > \varepsilon\right\} = 0.$$

Proof of Lemma 12: For $i = 1, \ldots, k_0$, by stationarity we have

$$\begin{aligned}\text{var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_{t,m,i}\right) &= \frac{1}{n} \sum_{t,s=1}^{n} \text{Cov}(Z_{t,m,i}, Z_{s,m,i}) \\ &= \frac{1}{n} \sum_{|h|<n} (n - |h|)\text{Cov}(Z_{t,m,i}, Z_{t-h,m,i}) \\ &\leq \sum_{h=-\infty}^{\infty} \left|\text{Cov}(Z_{t,m,i}, Z_{t-h,m,i})\right|.\end{aligned}$$

[2] DAVYDOV, Y. A. (1968) Convergence of Distributions Generated by Stationary Stochastic Processes. *Theory of Probability and Applications* 13, 691–696.

Because $\|d_\ell\| \leq K\rho^\ell$, we have

$$|Z_{t,m,i}| \leq K \sum_{\ell=m+1}^{\infty} \rho^\ell \|e_t\| \|e_{t-\ell}\|.$$

Using also $E\|e_t\|^4 < \infty$, it follows from the Hölder inequality that

$$\sup_h |\mathrm{Cov}(Z_{t,m,i}, Z_{t-h,m,i})| = \sup_h |EZ_{t,m,i}Z_{t-h,m,i}| \leq K\rho^m. \tag{38}$$

Let $h > 0$ such that $[h/2] > m$. Write

$$Z_{t,m,i} = Z^{h^-}_{t,m,i} + Z^{h^+}_{t,m,i},$$

where

$$Z^{h^-}_{t,m,i} = 2 \sum_{\ell=m+1}^{[h/2]} e'_{t-\ell} d'_\ell \Sigma^{-1}_{e0} e_t, \quad Z^{h^+}_{t,m,i} = 2 \sum_{\ell=[h/2]+1}^{\infty} e'_{t-\ell} d'_\ell \Sigma^{-1}_{e0} e_t.$$

Note that $Z^{h^-}_{t,m,i}$ belongs to the σ-field generated by $\{e_t, e_{t-1}, \ldots, e_{t-[h/2]}\}$ and that $Z_{t-h,m,i}$ belongs to the σ-field generated by $\{e_{t-h}, e_{t-h-1}, \ldots\}$. Note also that $\alpha_e(\cdot) = \alpha_\epsilon(\cdot)$ and that, by **A7**, $E|Z^{h^-}_{t,m,i}|^{2+\nu} < \infty$ and $E|Z_{t-h,m,i}|^{2+\nu} < \infty$. Lemma 11 then entails that

$$\left|\mathrm{Cov}(Z^{h^-}_{t,m,i}, Z_{t-h,m,i})\right| \leq K\alpha_\epsilon^{\nu/(2+\nu)}([h/2]). \tag{39}$$

By the argument used to show (38), we also have

$$\left|\mathrm{Cov}(Z^{h^+}_{t,m,i}, Z_{t-h,m,i})\right| \leq K\rho^h\rho^m. \tag{40}$$

In view of (38), (39) and (40),

$$\sum_{h=0}^{\infty} |\mathrm{Cov}(Z_{t,m,i}, Z_{t-h,m,i})| \leq K\rho^m + K\sum_{h=m}^{\infty} \alpha_\epsilon^{\nu/(2+\nu)}(h) \to 0$$

as $m \to \infty$ by **A7**. We have the same bound for $h < 0$. The conclusion follows from the Markov inequality. □